AMENDMENTS TO THE SPECIFICATION

In the Specification

Please substitute the following amended paragraph(s) and/or section(s) (deleted matter is shown by strikethrough and added matter is shown by underlining):

Please insert the following at Page 7, line 26:

A system and method for automatic harvesting and qualification of dynamic database content is disclosed herein. In the embodiment disclosed in Figure 1, at least one database 2 is communicatively coupled with a computer system 20. The computer system 20 includes a communication means 22. In various embodiments, the communication means 22 is for communicating with at least one other computer including a database to facilitate the two-way flow of information between said computer system and at least one other computer. In various embodiments, the computer system 20 includes a storage means 24. In various embodiments, the storage means is for retention and recall of data communicated by or to the at least one other computer. In various embodiments, the storage means 24 is capable of storing documents 70. In various embodiments, the computer system 20 includes a processing means 26. In various embodiments, the processing means is for executing multiple software modules and performing comparisons between a user supplied query and a plurality of documents found in at least one other computer. In various embodiments, the system and method for automatic harvesting and qualification of dynamic databases includes an index 30. In various embodiments, the index 30 is communicatively coupled to the processing means 26. In various embodiments, the index 30 is for storing a plurality of pre-approved internet sites to be included in a series of queries. In various embodiments, the system and method for automatic harvesting and qualification of dynamic databases includes a configuration module 40. In various embodiments, the configuration module 40 is communicatively coupled to the processing means 26. In various

embodiments, the configuration module 40 is adapted for translating a generic query into site-specific dialects such that a single user defined query may be directed to multiple sites automatically. In various embodiments, the system and method for automatic harvesting and qualification of dynamic databases includes a selection module. In various embodiments, the selection module is adapted for characterizing a plurality of documents returned by at least one database of at least one other computer and associated with a user defined query. In various embodiments, the system and method for automatic harvesting and qualification of dynamic databases includes a results index 50. In various embodiments, the results index 50 is communicatively coupled to the processing means 26. In various embodiments, the results index 50 allows for rapid recovery of specific portions of any one of a plurality of documents characterized by the selection module. In various embodiments, the system and method for automatic harvesting and qualification of dynamic databases includes a generator module 60. In various embodiments, the generator module 60 is for automatically generating at least one results page for the user conveying information associated with any one of a plurality of documents associated with a query.

Page 8, line 19-line 24:

The system uses a first one of the parametric information lists is a candidate database list, which provides an extensive group of candidate databases to be considered 201. The candidate databases can extend into the tens of thousands to hundreds of thousands. For example, on the Internet today, it is estimated there may be perhaps on the order of 250,000 searchable dynamic databases.

Page 8, line 26- Page 9, line 4:

An initial page from each of the initial listing of databases is captured 301. ~~An initial page~~ The initial page presented by each candidate database is evaluated for relevance 303 to the specific domain and subclassification of information or subject area 202. Any database which is determined to not be relevant to the subject area is removed from consideration for that subject area 304. A number of the remaining databases are selected for further consideration. The specific number of databases selected may be limited by a user-defined parameter 302 (such as a database relevancy parameter), which establishes a minimum threshold of relevancy for any given subject area.

Page 9, line 6- Page 9, line 15:

Each of the selected databases may have a unique set of requirements for submitting queries and retrieving documents. In order to facilitate the efficient harvest of content, each of the selected databases is analyzed for these requirements and a configuration file is created. For each database, the configuration file may serve as a translator between a generic query established by the user and the unique requirements of each database. The configuration file provides the system with information for the proper submission of queries and retrieval of responses for each one of the selected databases 203.

Page 9, line 17 – Page 9, line 27:

Each of the selected and configured databases is then again evaluated for relevance to the subject area 204. A sample query from the subject area is submitted to each of the selected databases 305. Responsive pages or documents are then gathered from each of the databases 306. These responsive documents are evaluated for relevance to the subject area 307. Each of the databases is assigned a numerical score representing relevance to the subject area 308. An

aggregate score may be developed 309. Databases with a sufficiently high numerical score are then qualified for use in the subject area 310. A different collection of databases may be qualified for each subject area. The qualified databases are then used for the next major function: document acquisition.

Page 10, line 5 – Page 10, line 12:

A difference between the initial harvest and the query servicing modes occurs at this point in the overall process. In an initial harvest the responsive content is captured or downloaded from the qualified database 205, 207. In the query servicing mode, the central location is checked for the document before resorting to downloading the document from the source database. If the central location has a current copy of the document, the systems resources are not used to download a new copy from the source database.

Page 10, line 14 – Page 10, line 19:

The system next performs the major function of indexing the content 208 for facilitating searching of the content. Here again is a difference between the initial harvest and query servicing modes. The index is created for documents qualified after the initial harvest. The index is used to find content matching a query during the query servicing mode.

Page 10, line 21 – Page 10, line 31:

The system parses each piece of content into constituent words 402 for processing. The system then compares each of the words to a fourth one of the parametric list (such as a stop list) 401, 206. A stop list contains terms which have been determined not to add value to the index, and therefore these terms are not processed. Each word, which is not on the stop list, is then stemmed into its base prefix (such as a stem word) to facilitate efficient indexing. The words on

the stop list are eliminated 403. The location of each stem word in every piece of content is then recorded 404 in the index, such that a user can search for any term based upon its corresponding stem word throughout the entire collection of content or documents through the index.

Page 11, line 1 – Page 11, line 9:

A summary of each piece of content may be created 210 if a summary was not provided by the qualified database. The summary may provide a listing of keywords relevant to the subject area, or an extract of a particularly relevant portion of the piece of content. This is especially important for content taken from large databases of documents (such as, for example, patent databases) where summaries for each document are typically not provided or available.

Page 11, line 11 – Page 11, line 18:

As a final step in the indexing process, the system records a plurality of statistics associated with each piece of content 211. Illustratively, the plurality of statistics may include, but is not limited to: the title of the piece of content, the number of internal links in the piece of content, the number of external links in the piece of content, the number of terms in the piece of content, the length of the piece of content, the database which provided the piece of content, and whether the content was static or dynamic.

Page 11, line 28 – Page 12, line 18:

After all of the queries have been submitted to the qualified databases and the responsive content has been captured 205 and stored 207 in a central location, the system matches each piece of responsive content to the initial categorization structure. The initial categorization structure is a tree configuration with each domain being a first level of classification and each sub-classification being a branch depending from the first level of classification or another sub-

classification. After this match has been performed, the system filters the categorization structure. This filtering may include a check for duplicate documents matched to the same classification, limiting the number of documents matched to any one classification or sub-classification based on a user defined parameter (such as a population parameter), and limiting the number of classifications or sub-classifications to which any one piece of content may be matched, based on a user defined parameter (such as an occurrence parameter). Additionally, the system may use a second parametric listing (such as an exclusion list) and a third parametric listing (such as an inclusion list) to inhibit matches or restrict matches (respectively) based upon a predetermined listing of terms and database sources for each subject area. After the filtering is complete, a categorization file is created 209 which records the matches of the stored copies of the responsive content for each subject area.

Page 12, line 20 – Page 12, line 25:

Finally, the system generates pages facilitating the recall of any piece of content in associate with a user's query 212. The user may submit a query to the system. The system will then match the query to the harvested content and return a page providing a listing of each relevant piece of content in the collection, along with a summary of the piece of content.